

Cluster formation and evolution in networks of financial market indices

Leonidas Sandoval Junior

Inspere, Instituto de Ensino e Pesquisa

November 23, 2011

Abstract

Using data from world stock exchange indices prior to and during periods of global financial crises, clusters and networks of indices are built for different thresholds and diverse periods of time, so that it is then possible to analyze how clusters are formed according to correlations among indices and how they evolve in time, particularly during times of financial crises. Further analysis is made on the eigenvectors corresponding to the second highest eigenvalues of the correlation matrices, revealing a structure peculiar to markets that operate in different time zones.

1 Introduction

This work uses a distance measure based on Spearman's rank correlation in order to build three dimensional networks based on hierarchies. This approach makes it possible to show how clusters form at different levels of correlations, and how the formations of those networks evolve in time.

The data used are the time series of international stock exchange indices taken from many countries in the world so as to represent both evolved and developing economies in a variety of continents. The series were taken both prior to and during some of the severest financial crises of the past decades, namely the 1987 Black Monday, the 1997 Asian Financial Crisis, the 1998 Russian Crisis, the burst of the dot-com bubble in 2001, and the crisis after September, 11, 2001, so that different regimes of volatility are represented. The idea is to be able to study how clusters behave when going from periods of low volatility to more turbulent times. The Subprime Mortgage crisis of 2008 is analyzed in a companion article, along with the years ranging from 2007 to 2010.

The data consists on some of the benchmark indices of a diversity of stock exchanges around the world. The number goes from 16 indices (1986) to 79 indices (2001). The choice of indices was mainly based on availability of data, since some of the stock exchanges being studied didn't develop or had no recorded indices until quite recently. As there are differences between some of the days certain stock markets operate, some of the operation days had to be deleted and others duplicated. The rule was the following: when more than 30% of markets didn't operate on a certain day, that day was deleted. When that number was below 30%, we repeated the value of the index for the previous day for the markets that didn't open. Special care had to be given to markets whose weekends didn't correspond to the usual occidental days, like some Arab countries. For those markets, we adjusted the weekends in order to match those of the majority of markets. All that was done in order to certify that we minimized the number of days with missing data, so that we could measure the log-returns of two consecutive days of a market's operation whenever that was possible.

The correlation matrix of the time series of financial data encodes a large amount of information, and an even greater amount of noise. That information and noise must be filtered if one is to try to understand how the elements (in our case, indices) relate to each other and how that relation evolves in time. One of the most common filtering procedures is to represent those relations using a *Minimum Spanning Tree* [1]-[15], which is a graph containing all indices, connected by at least one edge, so that the sum of the edges is minimum, and which presents no loops. Another type of representation is that of *Maximally Planar Filtered Graphs* [16]-[21], which admits loops but must be representable in two dimensional graphs without crossings.

Yet another type of representation is obtained by establishing a number which defines how many connections (edges) are to be represented in a graph of the correlations between nodes. There is no limitation with respect to crossing of edges or to the formation of loops, and if the number is high enough, then one has a graph where all nodes are connected to one another. These are usually called *Asset Trees*, or *Asset Graphs* [22]-[29], for they are no trees in the network sense. Another way to build asset graphs is to establish a value (threshold) such that distances above it are not considered. This eliminates connections (edges) as well as indices (nodes), but also turn the diagrams more understandable by filtering both information and noise.

Some previous works using graphic representations of correlations between international assets (indices or otherwise) can be found in [5], [20], and [28].

In the present work, we shall build asset trees of indices based on threshold values for a distance measure between them that is based on the Spearman rank correlation among indices. By establishing increasing values for the thresholds, we shall be able to devise the formation of clusters among indices, and how they evolve in time. Details of the way that is done are given in Section 2.

A second topic is pursued in Section 3, where an analysis is made of the eigenvector corresponding to the second highest eigenvalue of the correlation matrix of each group of data being studied. It is well-known ([30] and references therein) that the eigenvector corresponding to the highest eigenvalue of a correlation matrix of financial assets corresponds to a “market mode” that resembles the common movement of the market as a whole, and that often the eigenvectors corresponding to the second and sometimes the third highest eigenvalues also carry some information on the internal structure of the market being studied. What is implied in Section 3 is that this internal structure reflects the fact that stock markets operate at different time zones, what is a characteristic that is unique to a global market.

2 Depiction and evolution of clusters

In this section, we use data from some key periods of time for financial markets in order to build clusters based on hierarchy. The procedure is to consider the daily log-return for each index, given by

$$R_t = \ln P_t - \ln P_{t-1} , \quad (1)$$

where P_t is the value of the index at day t and P_{t-1} is the value of the same index on day $t - 1$. The log-returns are then used in order to create a correlation matrix C based on Spearman’s rank correlation, and then a distance is defined as

$$d_{ij} = 1 - c_{ij} , \quad (2)$$

where d_{ij} is the distance between indices i and j and c_{ij} is the correlation between both indices.

Spearman rank correlation is being used instead of the usual Pearson correlation, for it better captures non-linear relations between indices, and the distance measure was chosen as a linear realization of the correlation. The distance measure satisfies all conditions for an Euclidean measure.

In order to represent as best as possible the true distances of indices in a graph, we used three-dimensional maps based on principal component analysis, that minimizes the differences between the true distances and the distances represented in the graph. Then, those three-dimensional graphs were used in order to represent networks based on threshold values of the distances between nodes (indices). By using simulations (1000 for each period) with randomized data, where the log returns of each index were randomly shifted so as to eliminate any temporal correlation between them, but maintain their probability frequency distributions, we established thresholds above which random noise starts to interfere severely with the connections among indices. Those thresholds are discussed for each of the semesters depicted in this article.

A compromise had to be established between considering a good amount of data so as to minimize noise in the measurements of correlation and considering a small enough time interval so as to capture the relations between indices in a certain instant of time. Choosing to consider intervals of the size of a semester, this compromise was achieved, mainly because the crises being considered tended to happen on the second half of each year.

2.1 1987 - Black Monday

The first asset graphs to be built are based on data for the years 1986 and 1987, the year that preceded and the one that witnessed the crisis known as Black Monday, whose peak occurred on Monday, October 19, 1987. Not yet completely explained, the crisis made stock markets worldwide drop up to 45% in less than two weeks, and represented the worst financial crisis since 1929.

The networks for the two semesters of 1986 were built using the indices of 16 markets: S&P 500 from the New York Stock Exchange (S&P), and Nasdaq (Nasd), both from the USA, S&P TSX from Canada (Cana), Ibovespa from Brazil (Braz), FTSE 100 from the United Kingdom (UK), DAX from Germany (Germ) - West Germany at the time - ATX from Austria (Autr), AEX from the Netherlands (Neth), SENSEX from India (Indi), Colombo All Share from Sri Lanka (SrLa), Nikkei 25 from Japan (Japa), Hang Seng from Hong Kong (HoKo), TAIEX from Taiwan (Taiw), Kospi from South Korea (SoKo), Kuala Lumpur Composite from Malaysia (Mala), and JCI from Indonesia (Indo). The number of countries is small, mainly due to lack of data, but offers a relatively wide variety of nations and cultures in three continents. The networks for the two semesters of 1987 were built using 23 indices, adding to the ones of 1986 indices ISEQ from Ireland (Irel), OMX from Sweden (Swed), OMX Helsinki from Finland (Finl), IBEX 35 from Spain (Spai), ASE General Index from Greece (Gree), and PSEi from the Philippines (Phil).

2.1.1 First semester, 1986

Figure 1 shows the three dimensional view of the asset trees for the first semester of 1986, with threshold ranging from $T = 0.3$ to $T = 0.8$. Since we are using only the indices that have connections below $T = 0.8$, not all indices are displayed in the following graphs.

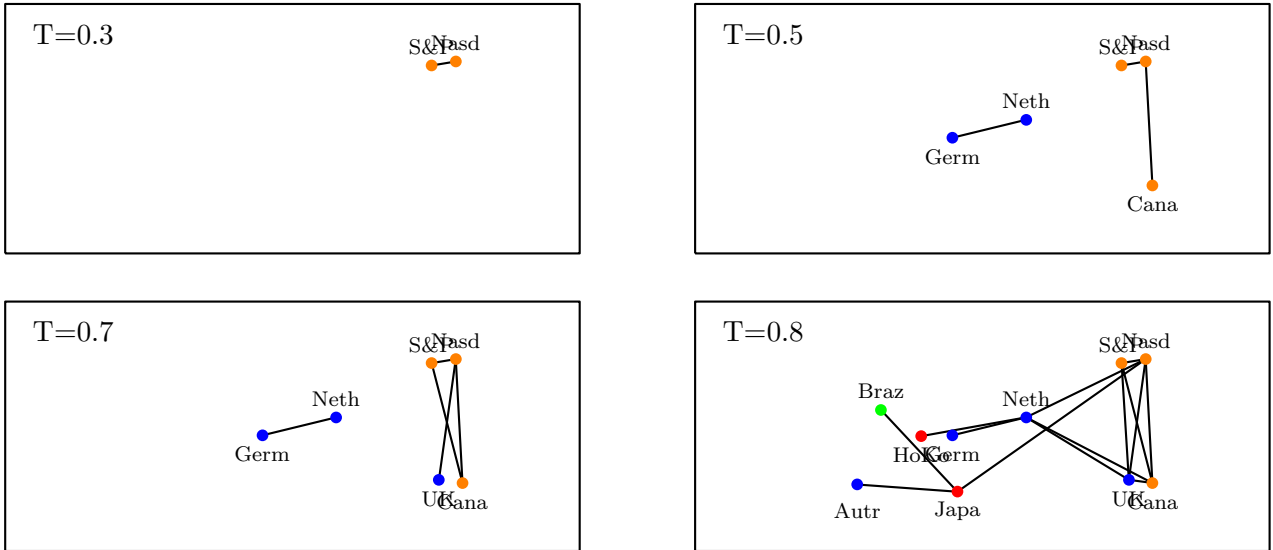


Fig. 1. Three dimensional view of the asset trees for the first semester of 1986, with threshold ranging from $T = 0.3$ to $T = 0.8$.

At $T = 0.3$, the first cluster is formed, with the connection between S&P and Nasdaq, both indices from the USA. For $T = 0.5$, Canada is added to the North American cluster and a new cluster is formed when Germany and Netherlands connect with each other. For $T = 0.6$, a connection is formed between S&P and Canada, strengthening the North American cluster. At $T = 0.7$, random noise starts to interfere with the results, but the connection formed between the UK and Nasdaq seems genuine. For $T = 0.8$, the two clusters merge, and Austria, Brazil, Japan, and Hong Kong join the new single cluster. Noise overwhelms the results for $T > 0.9$, and although many more connections are made for higher thresholds, they cannot be trusted. All the remaining indices connect with the resulting cluster within $T = 0.9$ and $T = 1.0$, and the last connections occur for $T = 1.3$.

2.1.2 Second semester, 1986

Figure 2 shows the three dimensional view of the asset trees for the second semester of 1986, with threshold ranging from $T = 0.3$ to $T = 0.8$.

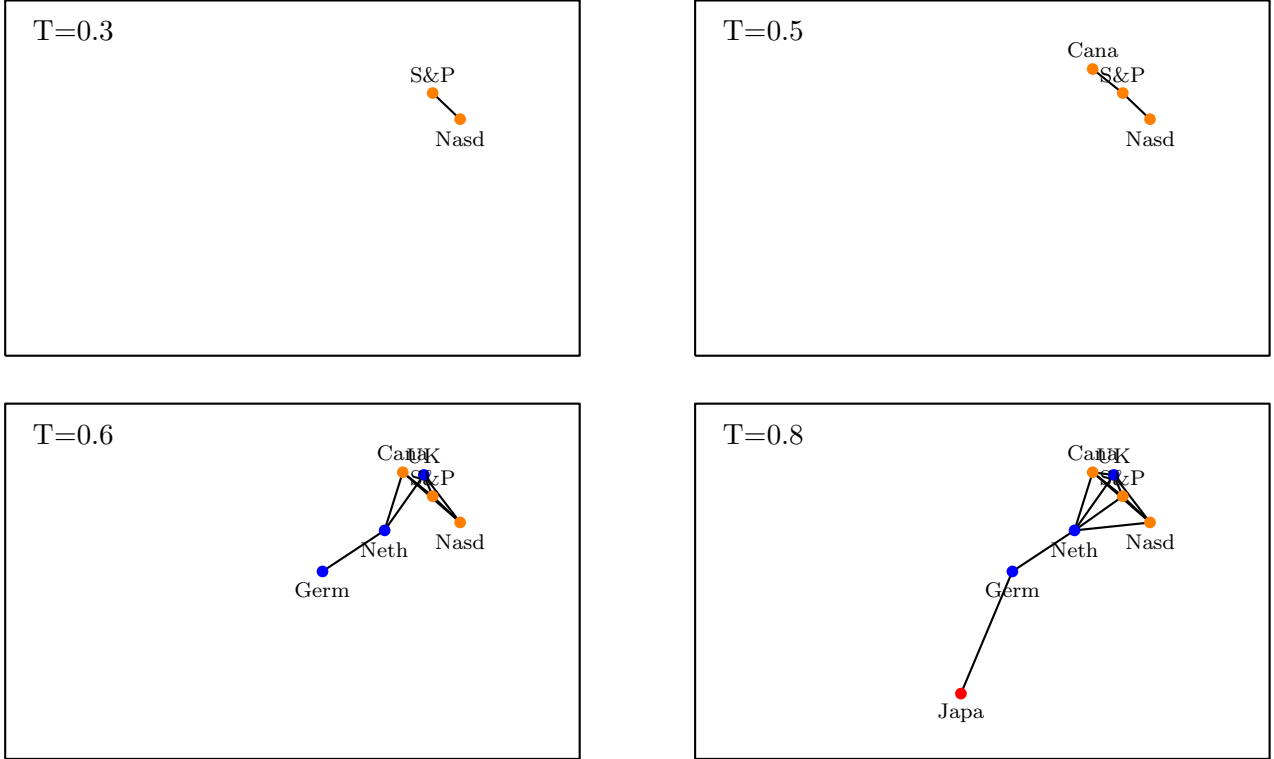


Fig. 2. Three dimensional view of the asset trees for the second semester of 1986, with threshold ranging from $T = 0.3$ to $T = 0.8$.

The first connection is formed at $T = 0.3$, between S&P and Nasdaq. This cluster grows only at $T = 0.5$, with the addition of Canada. At $T = 0.6$, the UK joins the North American cluster, which also connects with the Netherlands and, through it, with Germany. For $T = 0.7$, Netherlands establishes itself as a major hub, and Japan joins, via Germany, at $T = 0.8$. Random noise starts to have more effects at $T = 0.7$, and by $T = 0.9$, it becomes strong, so the many connections that are formed above $T = 0.9$ cannot be trusted. The last connections occur at $T = 1.3$.

2.1.3 First semester, 1987

Figure 3 shows the three dimensional view of the asset trees for the first semester of 1987, with threshold ranging from $T = 0.3$ to $T = 0.8$.

The first connection, again between S&P and Nasdaq, forms at $T = 0.3$. At $T = 0.4$, Canada joins the emerging North American cluster. For $T = 0.5$, a new cluster is formed, between Germany and Netherlands. At $T = 0.6$, a connection is formed between Canada and S&P, and for $T = 0.7$, the UK connects with the North American cluster through Nasdaq. For $T = 0.8$, random noise starts having strong effects. We have connections of the UK with Ireland and Hong Kong, and more connections among the members of the emerging European cluster, now comprised of Germany, Netherlands, Sweden, Finland, and Spain. South Korea connects with Ireland, and Sri Lanka connects with Malaysia. The last connections are probably due to random noise.

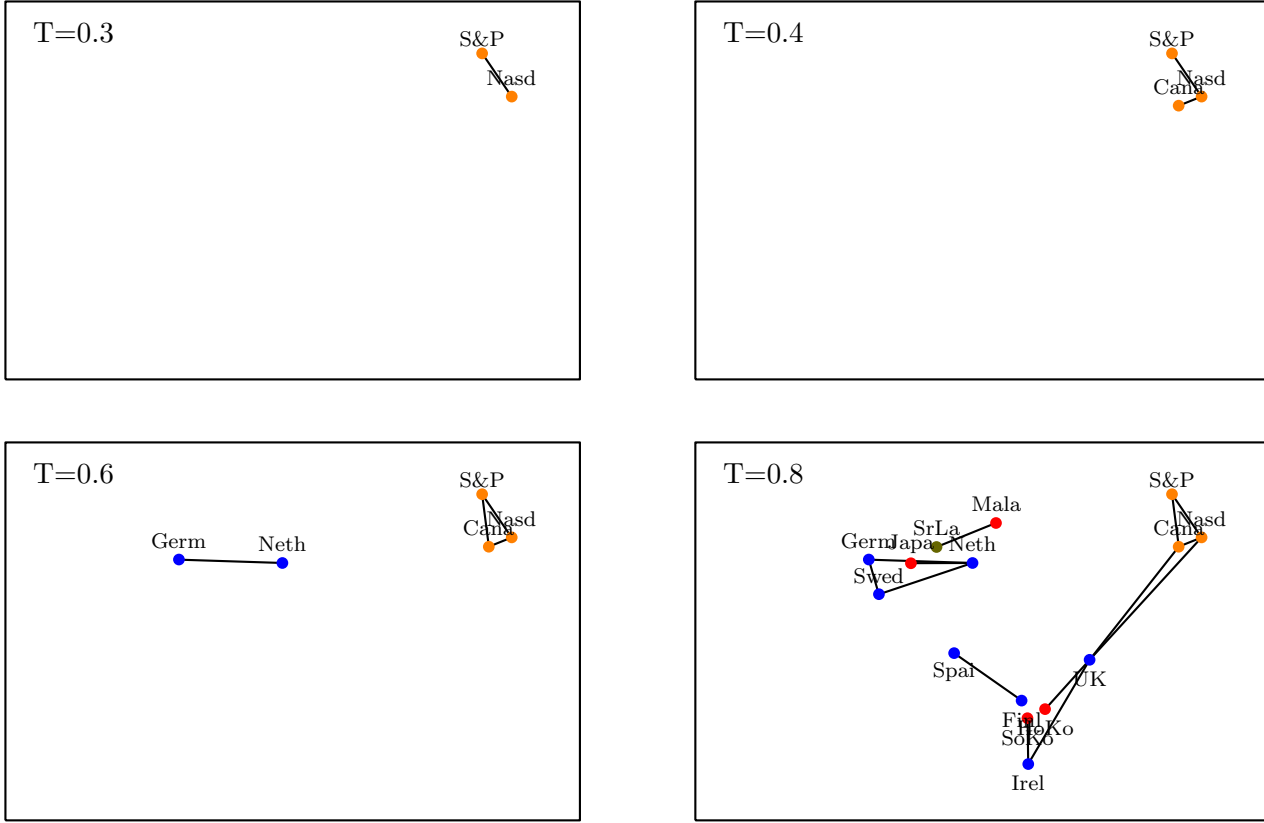
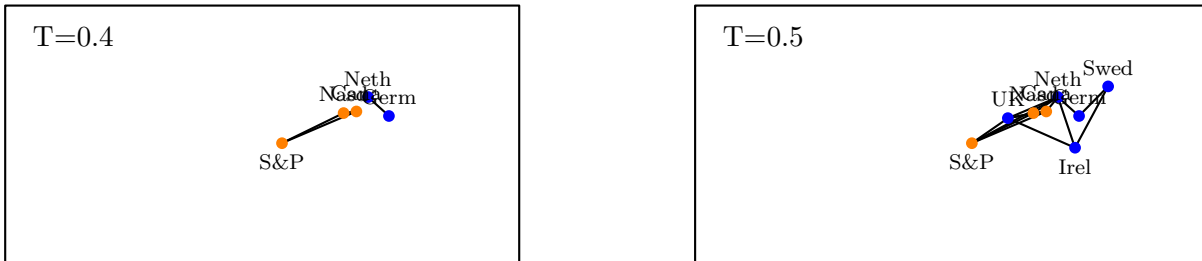


Fig. 3. Three dimensional view of the asset trees for the first semester of 1987, with threshold ranging from $T = 0.3$ to $T = 0.8$.

2.1.4 Second semester, 1987

Figure 4 shows the three dimensional view of the asset trees for the second semester of 1987, with threshold ranging from $T = 0.4$ to $T = 0.7$.

At $T = 0.3$, we have a cluster formed by the North American indices, S&P, Nasdaq, and Canada. For $T = 0.4$, the North American cluster gets fully connected, and a new cluster, composed by Germany and Netherlands, is formed. At $T = 0.5$, many connections are established, the UK and Netherlands with the North American cluster, between the UK and Ireland, and between Sweden and other European indices. At $T = 0.6$, Germany connects fully with the North American cluster, and Austria and Spain join the European cluster, while Japan connects with Sweden. Random noise starts to get much more influential for $T = 0.7$. We now have connections of India, Hong Kong, and Malaysia with the North American and European clusters. More connections are formed for higher thresholds, until all indices are connected for $T = 1.3$. Note that the size of the clusters shrank for the second semester of 1987, a consequence of the increase in correlation among the indices.



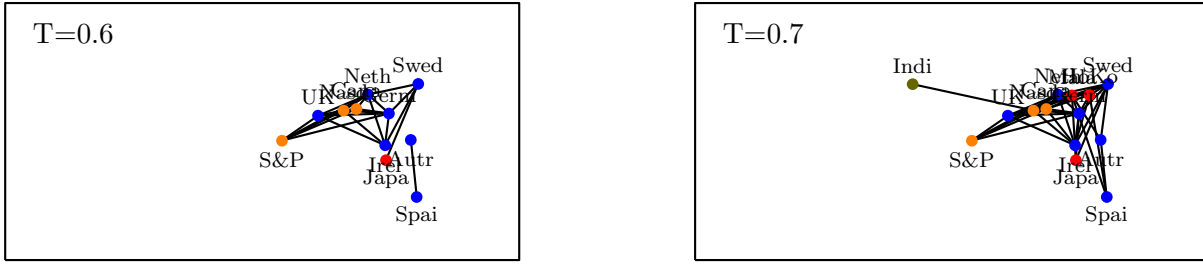


Fig. 4. Three dimensional view of the asset trees for the second semester of 1987, with threshold ranging from $T = 0.4$ to $T = 0.7$.

2.2 1997 and 1998 - Asian Financial Crisis and Russian Crisis

We now jump some years ahead to the next two crises we are going to analyze. The first one is the Asian Financial Crisis of 1997, which began with the devaluation of the Thai currency and spread to other Pacific Asian markets. The second was almost a consequence of the first one, as prices of commodities fell worldwide, affecting the Russian economy with particular acuteness. The networks for 1997 are built using 57 indices, adding the indices IPC from Mexico (Mexi), BCT Corp from Costa Rica (CoRi), Bermuda SX Index from Bermuda (Bermu), Jamaica SX from Jamaica (Jama), Merval from Argentina (Arge), IPSA from Chile (Chil), IBC from Venezuela (Vene), IGBVL from Peru (Peru), CAC 40 from France (Fran), SMI from Switzerland (Swit), BEL 20 from Belgium (Belg), OMX Copenhagen 20 from Denmark (Denm), OBX from Norway (Norw), OMX Iceland from Iceland (Icel), PSI 20 from Portugal (Port), PX (or PX50) from the Czech Republic (CzRe), SAX from Slovakia (Slok), Budapest SX from Hungary (Hung), WIG from Poland (Pola), OMXT from Estonia (Esto), ISE National 100 from Turkey (Turk), Tel Aviv 25 from Israel (Isra), BLOM from Lebanon (Leba), TASI from Saudi Arabia (SaAr), MSM 30 from Oman (Ohma), Karachi 100 from Pakistan (Paki), Shanghai SE Composite from China (Chin), SET from Thailand (Thai), S&P/ASX 200 from Australia (Aust), CFG25 from Morocco (Moro), Ghana All Share from Ghana (Ghan), NSE 20 from Kenya (Keny), FTSE/JSE Africa All Share from South Africa (SoAf), and SEMDEX from Mauritius (Maur). For 1997, we have no data about Russia, which is added as the index MICEX from Russia (Russ) to the data used for building networks for 1998 (which then has 58 indices).

2.2.1 First semester, 1997

Figure 5 shows the three dimensional view of the asset trees for the first semester of 1997, with threshold ranging from $T = 0.4$ to $T = 0.7$.

At threshold $T = 0.3$, the only connection is between S&P and Nasdaq. For $T = 0.4$, Canada joins the North American cluster, and an European cluster appears, formed by France, Germany, Switzerland, Belgium, Netherlands, Sweden, Finland, and Norway. At $T = 0.5$, nothing changes in the North American cluster, but the European cluster becomes denser, with more connections formed among its indices. The UK, Ireland, Austria, Denmark, Spain, and Portugal join that cluster. Something to notice is that there are strong ties formed among the Scandinavian indices, and among the Central European ones. For $T = 0.6$, Brazil and Argentina connect with the now American cluster. There is also further consolidation in the European cluster. Strange connections are formed, like two connections of Peru, with Denmark and Portugal. Less strange are the connections of Australia with Germany, and of Hong Kong with Austria. Noise starts to become important at this threshold, so one should be careful not to consider all connections now as true information. At $T = 0.7$, noise becomes a strong effect, but we analyze this case, anyway, because it also has some important information. We now have connections between the American and the European clusters, formed via Canada and Chile. Mexico connects with Argentina, and thus to the American cluster. Poland joins the European cluster, and Hong Kong, Australia, and South Africa establish more connections with Europe (South Africa, for the first time). There is also a small, separate cluster, formed by Malaysia and Indonesia. Peru connects with Norway, and Iceland forms a cluster with Mauritius, the latter almost clearly an effect of random noise. More connections are formed for higher thresholds, but many of them are the effect of random connections, and are indistinguishable from connections containing true information.

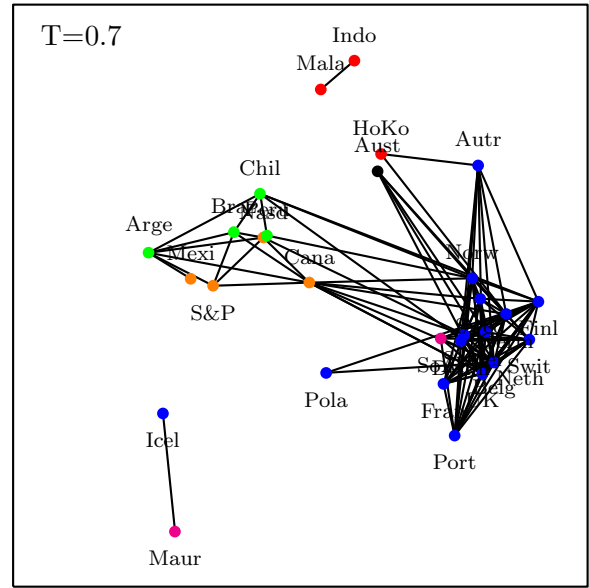
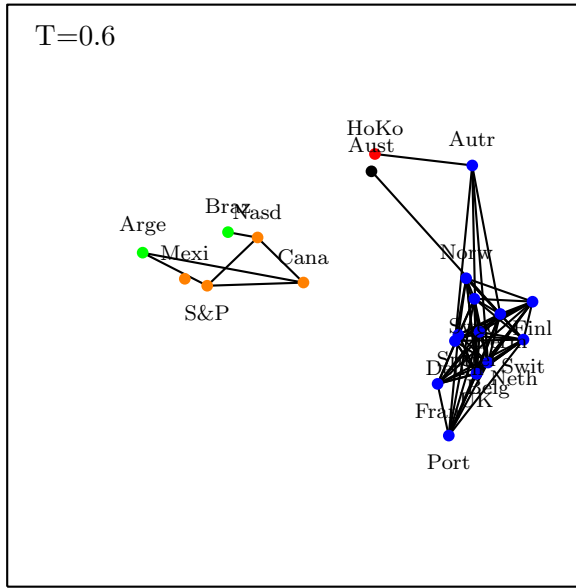
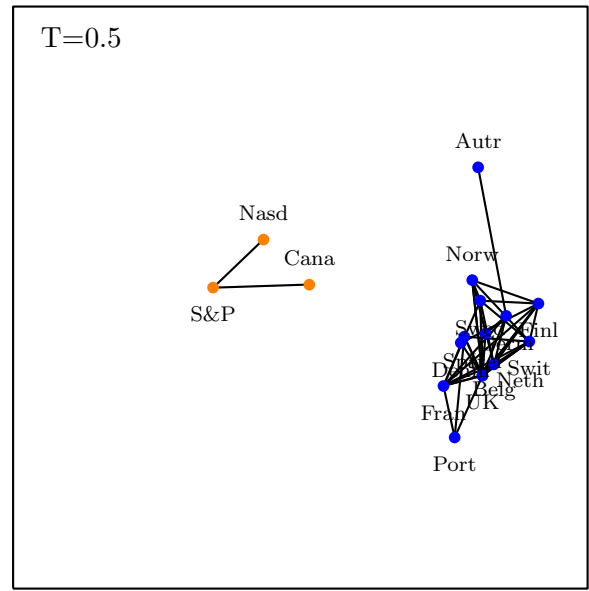
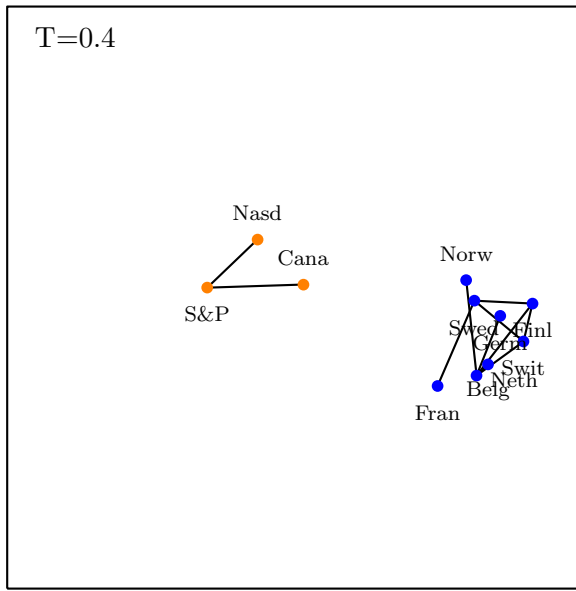
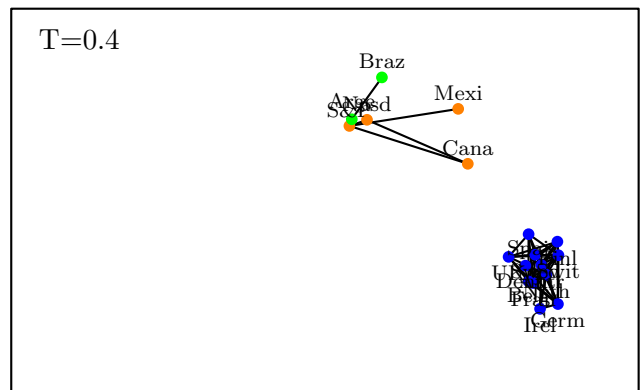
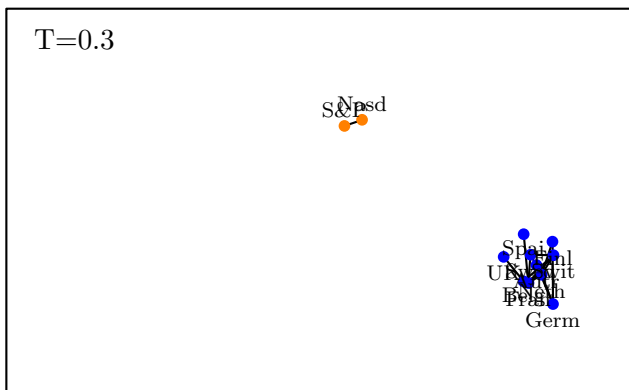


Fig. 5. Three dimensional view of the asset trees for the first semester of 1997, with threshold ranging from $T = 0.4$ to $T = 0.7$.

2.2.2 Second semester, 1997

Figure 6 shows the three dimensional view of the asset trees for the second semester of 1997, with threshold ranging from $T = 0.3$ to $T = 0.6$.



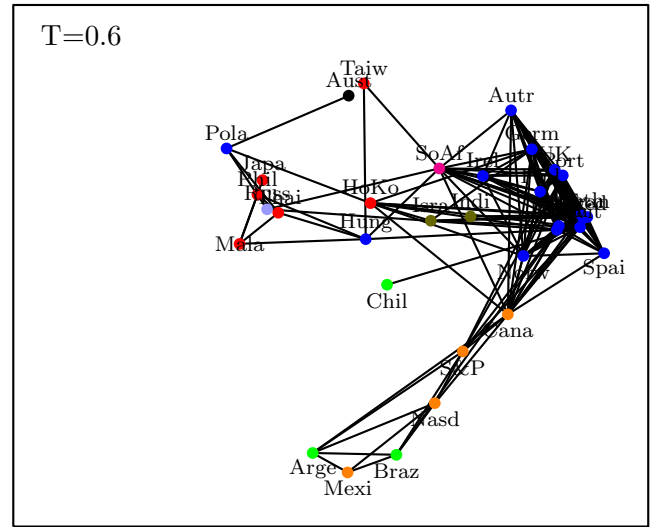
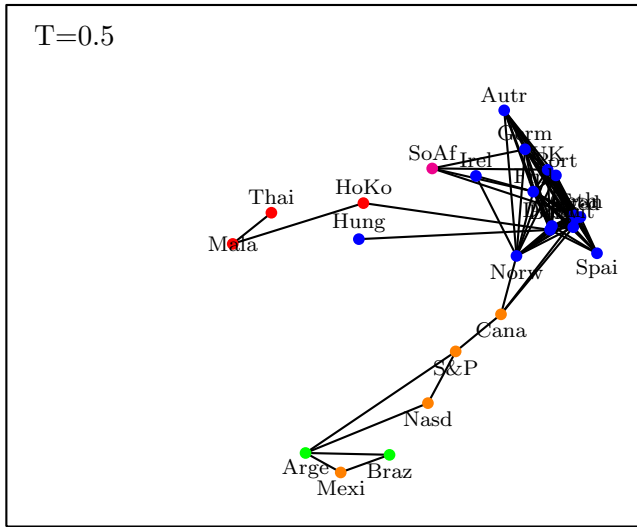
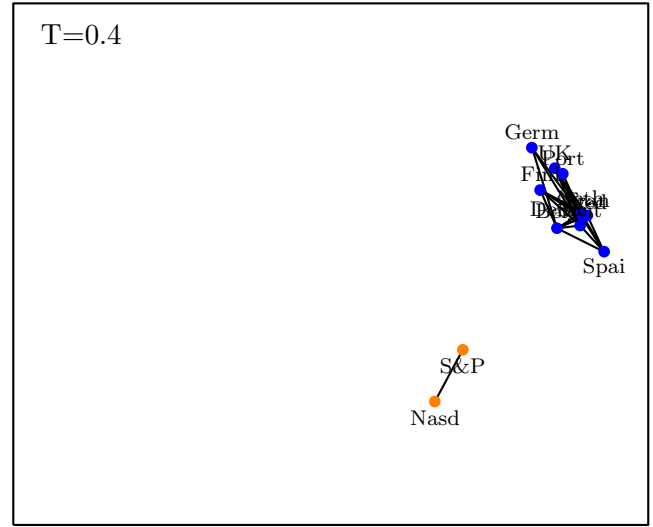
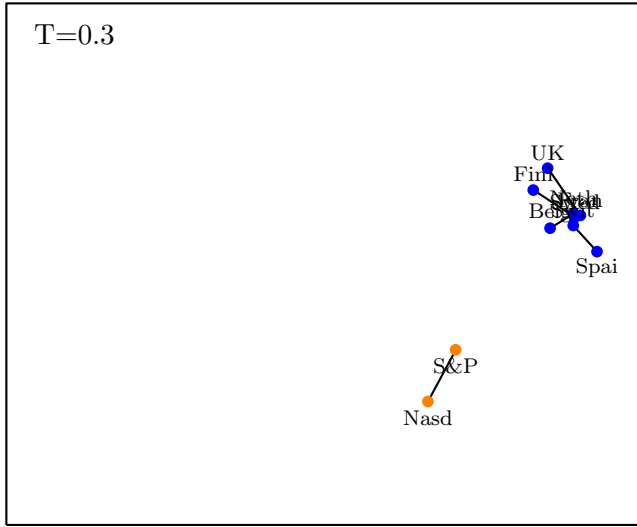
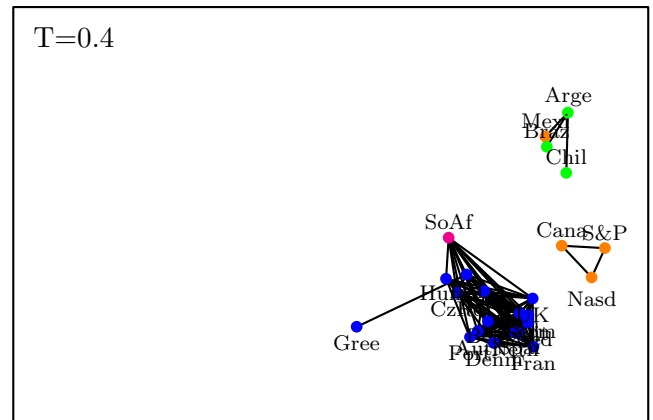
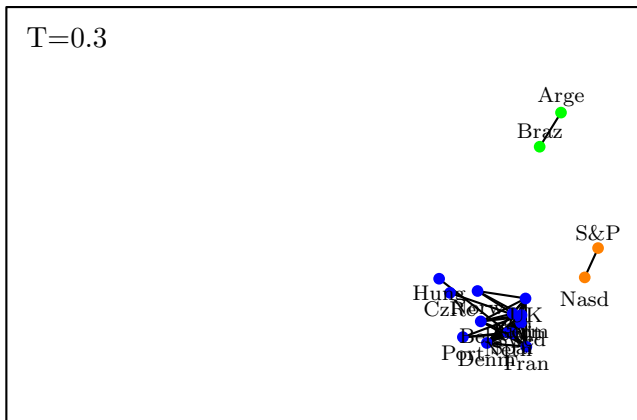


Fig. 7. Three dimensional view of the asset trees for the first semester of 1998, with threshold ranging from $T = 0.3$ to $T = 0.6$.

2.2.4 Second semester, 1998

Figure 8 shows the three dimensional view of the asset trees for the second semester of 1998, with threshold ranging from $T = 0.3$ to $T = 0.6$.



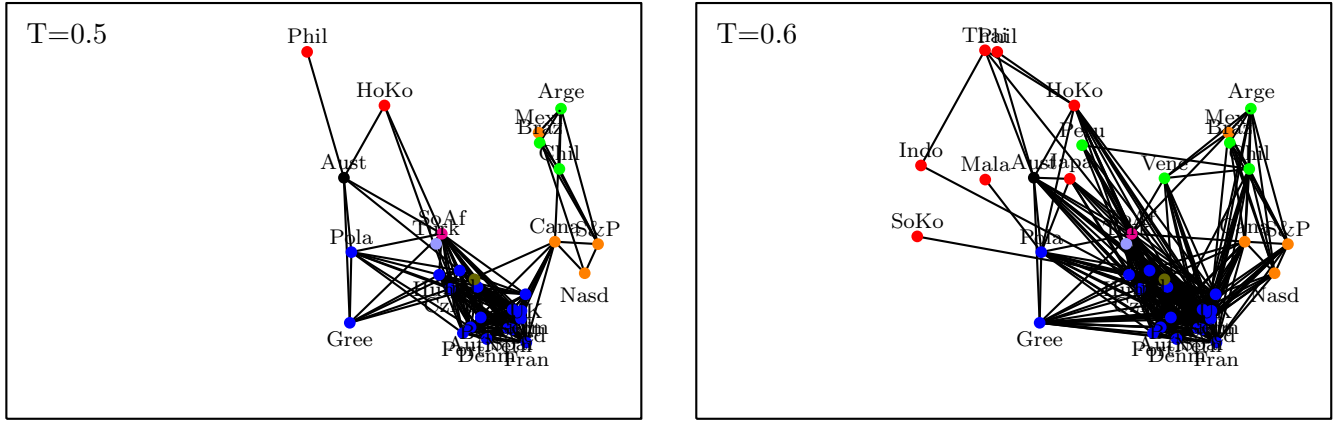


Fig. 8. Three dimensional view of the asset trees for the second semester of 1998, with threshold ranging from $T = 0.3$ to $T = 0.6$.

At $T = 0.2$, two clusters appear: the North American one, formed by S&P and Nasdaq, and the European cluster, comprised of the UK, France, Germany, Switzerland, Belgium, Netherlands, Sweden, Finland, and Spain. At $T = 0.3$, a third cluster, comprised of Brazil and Argentina, is formed. The European cluster strengthens its ties, and gains Denmark, Norway, Portugal, the Czech Republic, and Hungary. For $T = 0.4$, Canada joins the North American cluster, and Mexico and Chile join the South American one. The European cluster becomes even denser and absorbs Ireland, Austria, and Greece. South Africa makes connections with most of the European indices, effectively joining the European cluster. For $T = 0.5$, the North and South American clusters merge, and Canada makes many connections with Europe. Poland joins the European cluster, which now makes connections with Turkey, Israel, Hong Kong, and Australia. Hong Kong connects with South Africa, and Australia connects with Hong Kong, Philippines, and South Africa. At $T = 0.6$, there is a strong integration between the American and the European clusters, now with the addition of Venezuela and Peru. Turkey, Israel, Hong Kong, and Australia fully integrate with Europe, and Japan connects both with the European cluster and with a Pacific Asian one, comprised of Japan, Hong Kong, Thailand, Indonesia, and Australia. South Korea and Malaysia connect with Europe and not with the Pacific Asian cluster. For $T > 0.6$, the connections between clusters strengthen, and the Pacific Asian cluster becomes more self-connected, but noise becomes strong from here onwards, and so connections cannot be trusted anymore. The last connections occur at $T = 1.3$.

2.3 2001 - Burst of the dot-com Bubble and September 11

Two very distinct crises arose in 2001: the first was the burst of a bubble due to the overvaluing of the new internet-based companies, and the second was the fear caused by the severest terrorist attack a country has ever faced, when the USA was targeted by commercial airplanes overtaken by terrorists. Both crises were very different in origin and in duration, and their analysis might shed some light on how crises may develop.

For the networks concerning the year 2000, we add now the indices Bolsa de Panama General from Panama (Pana), FTSE MIB (or MIB-30) from Italy (Ital), Malta SX from Malta (Malt), LUXX from Luxembourg (Luxe), BET from Romania (Roma), OMXR from Latvia (Latv), OMXV from Lithuania (Lith), PFTS from Ukraine (Ukra), Al Quds from Palestine (Pale), ASE General from Jordan (Jorda), QE (or DSM 20) from Qatar (Qata), MSE Top 20 from Mongolia (Mong), Straits Times from Singapore (Sing), TUNINDEX from Tunisia (Tuni), EGX 30 from Egypt (Egyp), and Nigeria SX All Share from Nigeria (Nige), for a total of 74 indices. For the networks of 2001, we also add the indices SOFIX from Bulgaria (Bulga), KASE from Kazakhstan (Kaza), VN-Index from Vietnam (Viet), NZX 50 from New Zealand (NeZe), and Gaborone from Botswana (Bots), for a total of 79 indices.

2.3.1 First semester, 2000

Figure 9 shows the three dimensional view of the asset trees for the first semester of 2000, with threshold ranging from $T = 0.3$ to $T = 0.6$.

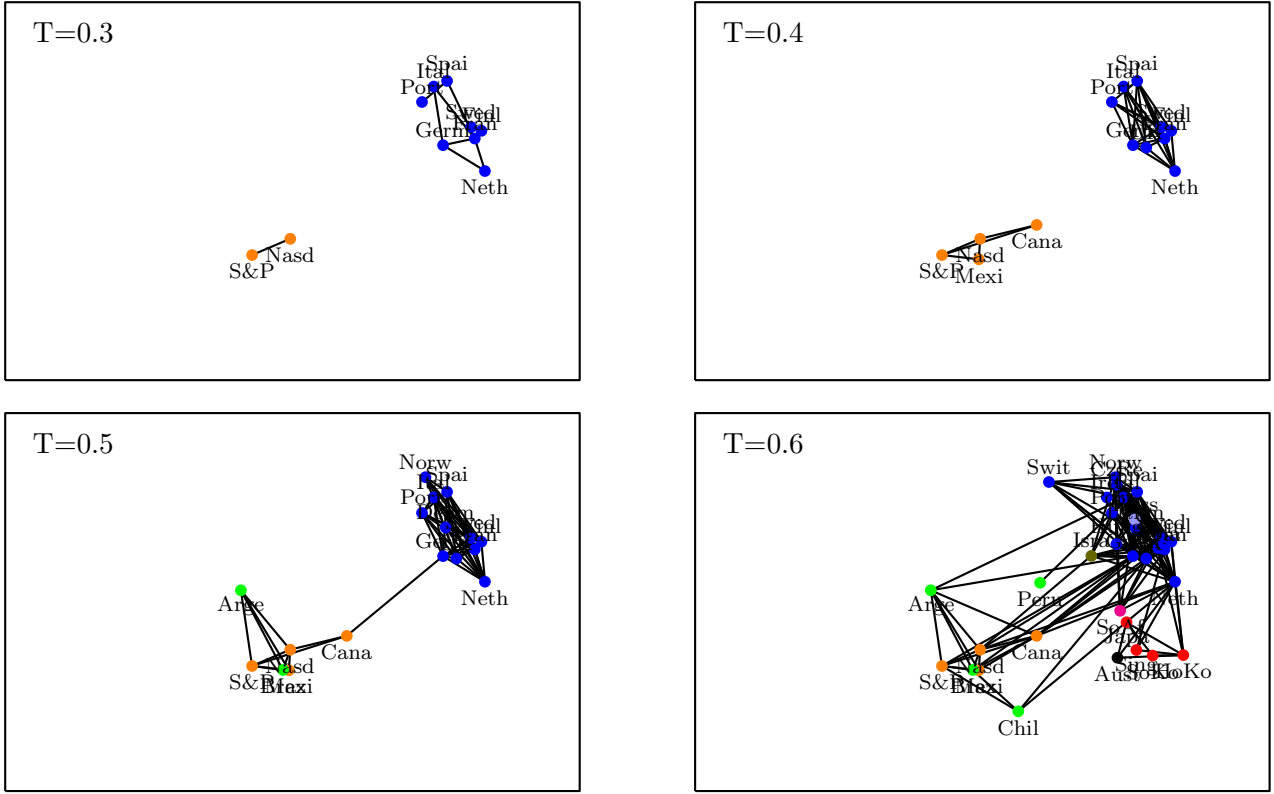


Fig. 9. Three dimensional view of the asset trees for the first semester of 2000, with threshold ranging from $T = 0.3$ to $T = 0.6$.

For $T = 0.2$, the only connection is between S&P and Nasdaq. For $T = 0.3$, a second cluster, comprised of European indices, is formed, with France, Germany, Italy, Netherlands, Sweden, Finland, Spain, and Portugal. At $T = 0.4$, the North American cluster is fully connected, with the addition of Canada and Mexico. The UK joins the European cluster, which also becomes denser, with the establishment of more connections between its members. For $T = 0.5$, Brazil and Argentina join the American cluster, and also connect with one another. Canada connects with Germany, in the European cluster, which now grows even denser, and has Denmark added to it. At $T = 0.6$, many more connections are formed, inside clusters and between clusters. Israel connects with both America and Europe, Hong Kong, Singapore, Australia and South Africa connect with Europe. Switzerland, Norway, the Czech Republic, Hungary, and Russia also join the European cluster. A Pacific Asian cluster is formed, comprised of Japan, Hong Kong, South Korea, and Singapore, strongly connected with Australia, and lightly connected with Europe. For larger thresholds, noise starts to dominate, and many connections are made, some of them apparently at random. At $T = 1.3$, all indices are connected, with the last connection being between Jamaica and Lebanon.

2.3.2 Second semester, 2000

Figure 10 shows the three dimensional view of the asset trees for the second semester of 2000, with threshold ranging from $T = 0.3$ to $T = 0.6$.

Once more, at $T = 0.2$, a single connection is formed between S&P and Nasdaq. For $T = 0.3$, we also have an European cluster comprised of the UK, France, Germany, Italy, Netherlands, Sweden, Finland, and Spain. At $T = 0.4$, the North American cluster, now comprised of S&P, Nasdaq, Canada, and Mexico, joins the European cluster via Germany. The European cluster has now only the addition of Portugal, but it becomes denser, with more connections formed among its members. At $T = 0.5$, Argentina joins the American

cluster, which makes stronger connections with Europe. Switzerland, Norway, and Russia join the European cluster. At $T = 0.6$, a myriad of connections is made, with Brazil and Chile joining the American cluster, and Ireland, Austria, Belgium, the Czech Republic, Hungary, and Poland joining the European Cluster. The connections between both clusters grow stronger, and Israel, Hong Kong, South Korea, Singapore, Australia, and South Africa establish connections with some or both clusters. There is also a new cluster, of Pacific Asian indices, comprised of Japan, Hong Kong, South Korea, and Singapore, with connections to Australia and to South Africa. At larger thresholds, new connections are made, subject to a lot of random noise, and the last connection is formed at $T = 1.4$, between Ireland and Lebanon.

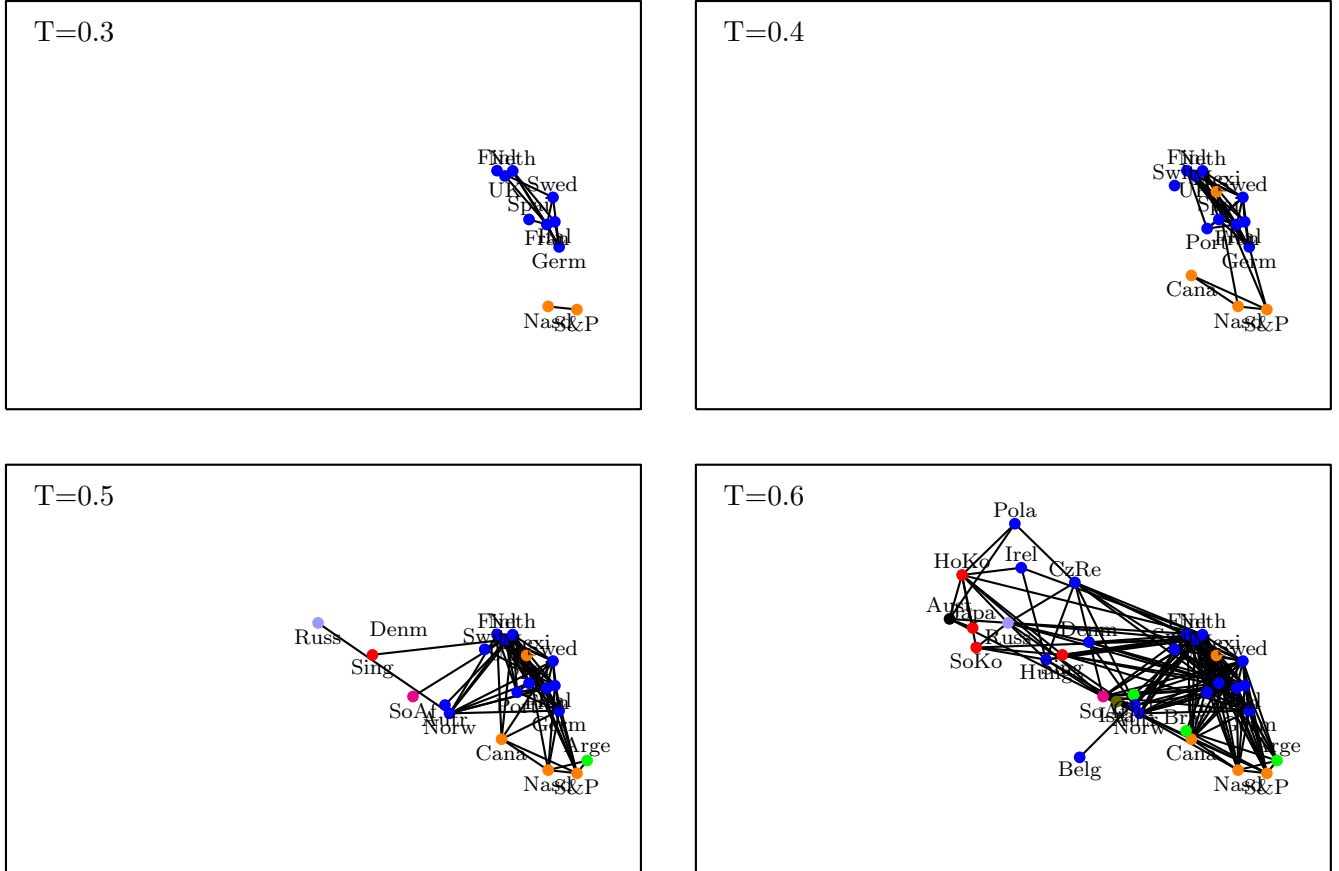


Fig. 10. Three dimensional view of the asset trees for the second semester of 2000, with threshold ranging from $T = 0.3$ to $T = 0.6$.

2.3.3 First semester, 2001

Figure 11 shows the three dimensional view of the asset trees for the first semester of 2001, with threshold ranging from $T = 0.3$ to $T = 0.6$.

At $T = 0.2$, there are two small clusters: one formed by *S&P* and Nasdaq, and the other formed by France, Italy, and Netherlands. At $T = 0.3$, Canada joins the North American cluster, and the UK, Germany, Sweden, and Spain join the European cluster. At $T = 0.4$, the European cluster is more interwoven, and have the addition of Switzerland, Belgium, Finland, Norway, and Portugal. At $T = 0.5$, the North American cluster is fully formed, with the addition of Mexico, and a South American cluster forms, composed of Brazil and Argentina. Europe becomes even more densely connected, and receives the addition of Ireland, Denmark, the Czech Republic, and Hungary, the last two considered as part of Eastern Europe. We also have Israel and South Africa connecting with Europe. A new cluster is born, made of Pacific Asian countries, namely Japan, Hong Kong, and South Korea. The North American and European clusters start to get connected. At $T = 0.6$, North and South America are integrated with each other and with Europe. Newcommers are Chile to South

America, Austria, Luxembourg, Greece, Poland, and Russia to the European cluster, Singapore to the Pacific Asian cluster, and New Zealand, which also connects with the Pacific Asian cluster. Through Singapore and Hong Kong, the Pacific Asian cluster connects with the European one. For $T > 0.6$, we see further integration and more markets joining the fold, but we then enter a region where noise dominates, and connections cannot be fully trusted. By $T = 1.3$, all indices are connected with one another.

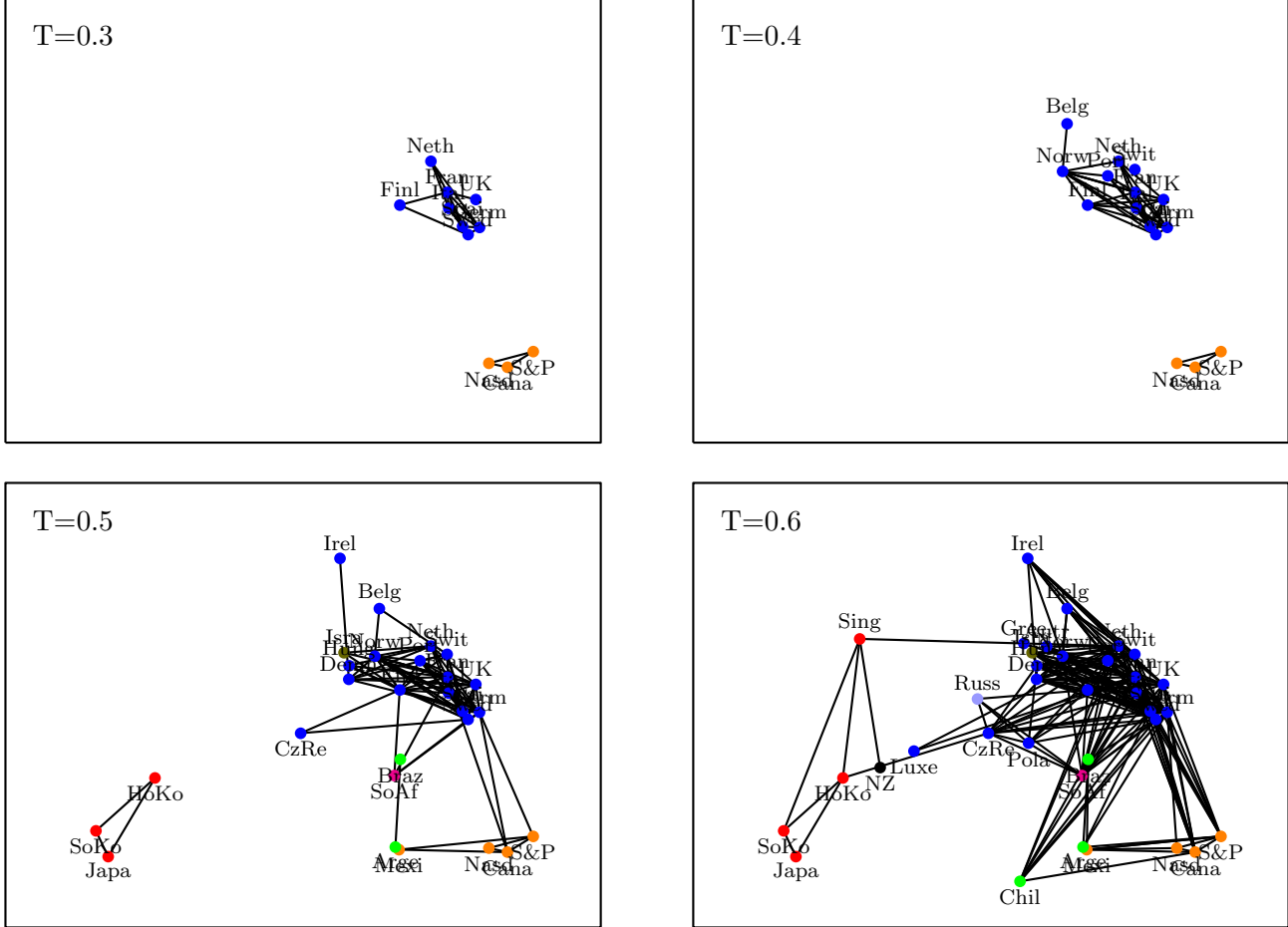


Fig. 11. Three dimensional view of the asset trees for the first semester of 2001, with threshold ranging from $T = 0.3$ to $T = 0.6$.

2.3.4 Second semester, 2001

Figure 12 shows the three dimensional view of the asset trees for the second semester of 2001, with threshold ranging from $T = 0.3$ to $T = 0.6$.

For the data concerning the second semester of 2001, for $T = 0.1$, there is already a connection between France and Netherlands. At $T = 0.2$, two small clusters are formed: S&P and Nasdaq (North American cluster) and an European cluster comprised of the UK, France, Germany, Switzerland, Italy, Netherlands, Sweden, and Spain. At $T = 0.3$, Canada joins the North American cluster, the European cluster becomes more densely knit, and Belgium and Finland join the fold. At $T = 0.4$, the North American cluster joins with the European one via Canada, and the European cluster receives the addition of Portugal.

At $T = 0.5$, there are many more connections formed: Mexico joins the North American cluster, and more connections are made between this cluster and the European one. Europe receives Ireland, Denmark, Norway, Greece, the Czech Republic, and Estonia. Strange connections form, like one between Sweden and Peru, and between Estonia and Hong Kong, what makes us think that random noise already has some effects here. At this same threshold, a Pacific Asian cluster is formed, already linked with Europe, comprised of Hong Kong, Taiwan, South Korea, and Singapore. Australia connects both with the Asian cluster and with the European

cluster. For $T = 0.6$, the threshold that we are considering as the limit to the region dominated by noise, Brazil connects with the North American cluster via Canada, and also connects itself with Chile, which establishes connections with Europe, while Peru connects with many European indices. More connections are formed between America and Europe, and the European cluster now adds to itself Luxembourg, Hungary, and Poland. Russia, Turkey, Israel, and South Africa connect with the European cluster, and Japan and India join the Pacific Asian one. At higher thresholds, an explosion of connections occur, some of them reinforcing already made clusters, some establishing more connections between them, and others connecting more indices to those clusters, many of them apparently at random.

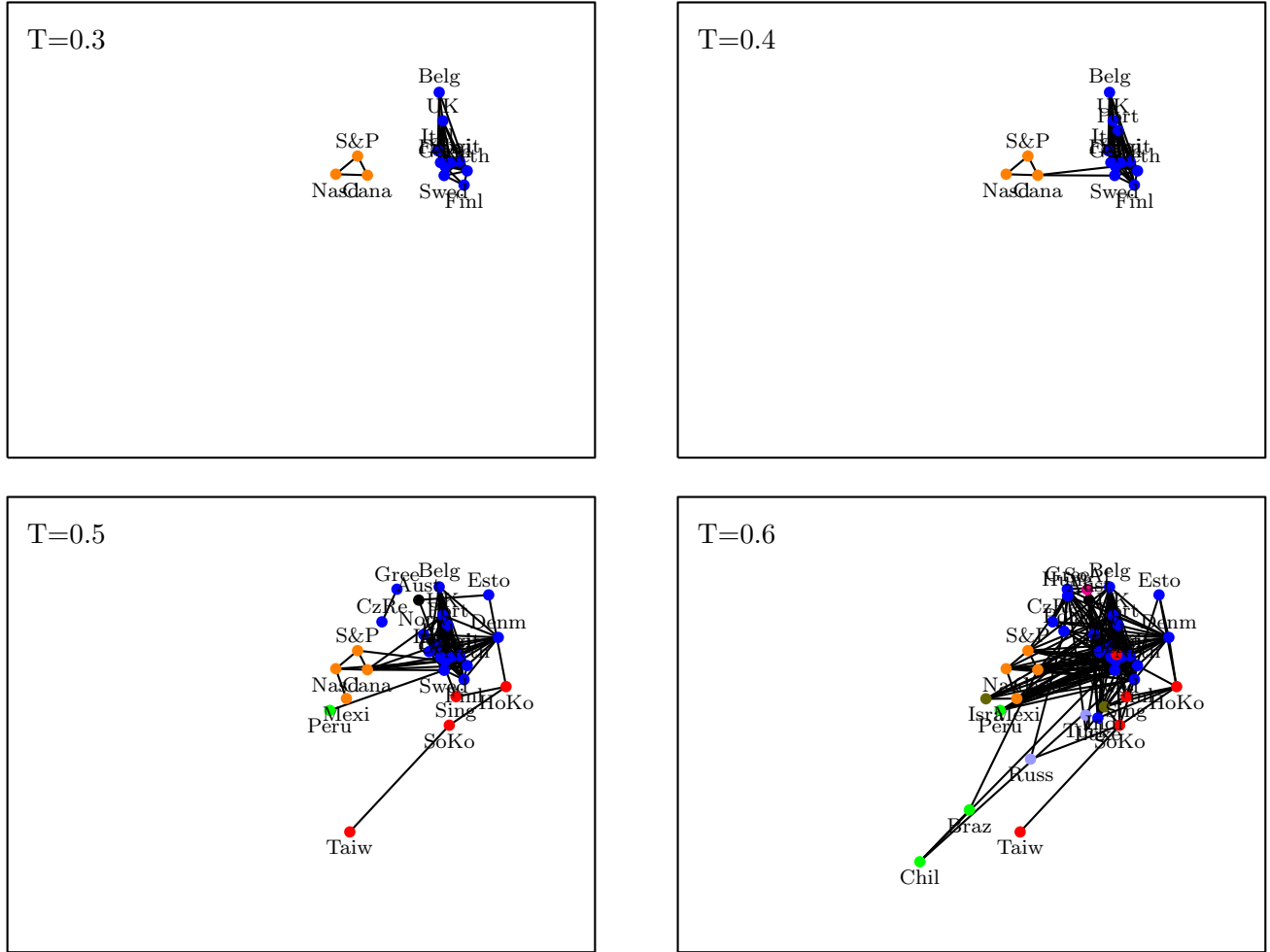


Fig. 12. Three dimensional view of the asset trees for the second semester of 2001, with threshold ranging from $T = 0.3$ to $T = 0.6$.

2.4 Overview

By considering all the time periods being studied, the first fact that stands out is the presence of two cluster throughout all periods: the American and the European clusters. The connections between the S&P 500 and Nasdaq are present at all times and at the smallest thresholds. What we call an European cluster also appears early on, but its core varies with time, beginning with the pair Germany-Netherlands and then evolving to a new core, consisting on the UK, France, Germany, Switzerland, Italy, Netherlands, Sweden, and Spain. It is interesting to notice that the UK, which was more connected to the American cluster in the 80's, moves to the European cluster later on.

At higher thresholds, the American cluster is joined by Canada and some South American indices, and the European cluster grows with the addition of some more indices of Scandinavian countries and of other Western European indices. As the thresholds grow, Europe is joined by Eastern European countries and only then

we witness the formation of a Pacific Asian cluster, which starts to solidify after 1997, which coincides with the Asian Financial Crisis. South Africa and, in a lesser proportion, Israel, connect with Europe rather than with their neighbouring countries. Australia and New Zealand are more connected with Pacific Asia than with Europe. Indices from the Caribbean, of most islands, of the majority of Africa, and of the Arab countries, connect only at much higher values of the threshold, where noise reigns.

What is also clear from the graphics is that the financial indices tend to group according to the geographic proximity of their countries, and also according to cultural similarities, as are the case of South Africa and Israel with Europe. Other striking feature is that the American and European clusters are fairly independent, and only connect at higher thresholds, when both cluster have already grown larger and denser. It is good to remind ourselves that networks built from correlation matrices are not directed networks, so that we can deduct no effects of causality between the indices using just the correlation matrix (an alternative would be parcial correlation [31]-[34]).

Also to be noticed is that networks shrink in size in times of crises, reflecting the growth in correlation between markets in those times [30].

3 Information in the second largest eigenvalue

It is well documented in the literature that the eigenvalues of the correlation matrix of the time series of assets may be used to identify part of the information carried by those time series separating it from noise. This is usually done using Random Matrix Theory [35], that is based on the analysis of the eigenvalue frequency distribution of a matrix which is obtained from the correlation between N time series of random numbers generated over a Gaussian frequency distribution with zero mean and standard deviation σ . If L is the number of elements of each time series, then, in the limit $L \rightarrow \infty$ and $N \rightarrow \infty$ such that $Q = L/N$ is constant, finite and greater than one, then the probability distribution function of the eigenvalues of such a matrix is given by the expression

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \quad (3)$$

where

$$\lambda_- = \sigma^2 \left(1 + \frac{1}{Q} - 2\sqrt{\frac{1}{Q}} \right), \quad \lambda_+ = \sigma^2 \left(1 + \frac{1}{Q} + 2\sqrt{\frac{1}{Q}} \right), \quad (4)$$

and λ is restricted to the interval $[\lambda_-, \lambda_+]$.

This probability distribution function is called a Marčenko-Pastur distribution [36], and it establishes limits for the eigenvalues generated from random data. So, in theory, any eigenvalue falling out of the interval $[\lambda_-, \lambda_+]$ has a good chance of representing true information about the system.

As stated before, the Marčenko-Pastur distribution is valid only for the limit of an infinite amount of available data, and also for random time series generated over a Gaussian. Since this is not the case for real time series of financial data, for they are finite and usually their probability frequency distributions are not Gaussian, an alternative would be to use randomized data for a collection of time series and then analyze the minimum and maximum values of eigenvalues found for the correlation matrix thus generated. The result is similar to the one obtained for theoretical distributions, but not quite the same.

In this section, we use the results obtained from randomized data in order to establish regions where noise can be a major concern and then analyze the eigenvectors of some eigenvalues that are outside those regions. Figures 13 to 15 show the eigenvalues (represented as lines) and the values associated with noise (shaded region) in the graphs for the years being studied in this article.

The first feature which is clearly visible is that the highest eigenvalue always stands out from the others. Each eigenvalue has a corresponding eigenvector with the same number of indices used in creating the correlation matrix. Each entry in an eigenvector corresponds to the participation of that index in the building of a “portfolio” of indices. It is well known that the eigenvectors of the highest eigenvalues are usually built in such a way that every index has a similar participation in such a portfolio, and that this particular combination emulates the general behavior of a world market index [30].

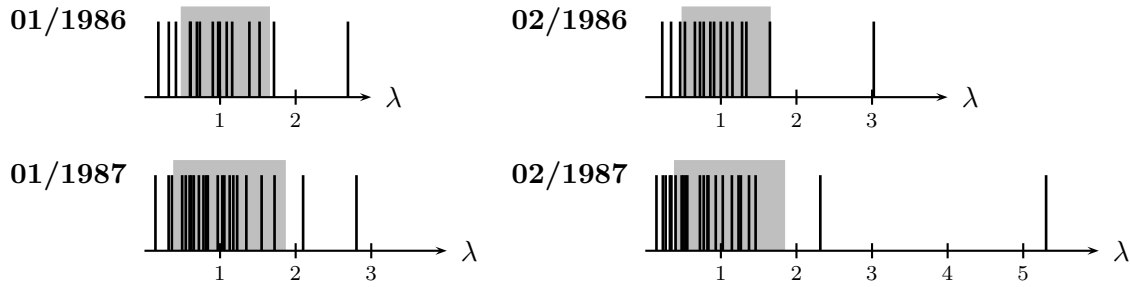


Fig. 13. Eigenvalues for the first and second semesters of 1986 and 1987 in order of magnitude. The shaded area corresponds to the eigenvalues predicted for randomized data.

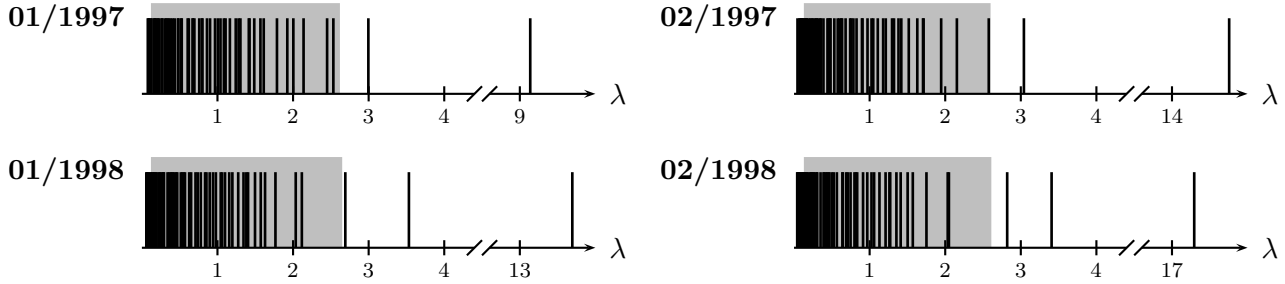


Fig. 14. Eigenvalues for the first and second semesters of 1997 and 1998 in order of magnitude. The shaded area corresponds to the eigenvalues predicted for randomized data.

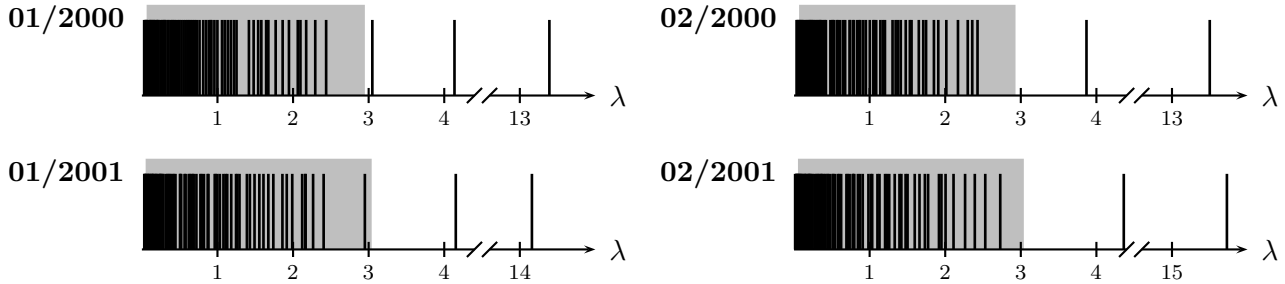


Fig. 15. Eigenvalues for the first and second semesters of 2000 and 2001 in order of magnitude. The shaded area corresponds to the eigenvalues predicted for randomized data.

To illustrate this, if one compares the log-returns of a portfolio consisting on the indices as expressed by the eigenvectors corresponding to the highest eigenvalues with the log-returns of the world index calculated by the MSCI (Morgan Stanley Capital International), then one obtains the correlations between those indices as stated in Table 1.

| Semester/Year | 01/1986 | 02/1986 | 01/1987 | 02/1987 | 01/1997 | 02/1997 | 01/1998 | 02/1998 |
|---------------|---------|---------|---------|---------|---------|---------|---------|---------|
| Correlation | 0.69 | 0.60 | 0.40 | 0.87 | 0.64 | 0.81 | 0.83 | 0.82 |

| Semester/Year | 01/2000 | 02/2000 | 01/2001 | 02/2001 |
|---------------|---------|---------|---------|---------|
| Correlation | 0.71 | 0.83 | 0.77 | 0.82 |

Table 1. Correlations between indices built on the eigenvector corresponding to the highest eigenvalue and the MSCI world index according to semester and year.

Although the correlation is weak for early years, probably due to the small number of indices that were considered for them, correlation is strong for subsequent years. So, the eigenvector corresponding to the highest eigenvalue is associated with a “market mode”, or the general oscillations common to all indices.

Now, we move our attention to the second highest eigenvalue. Figures 13 to 15 show that such eigenvalue is also detached from the noisy region, although less so than the highest eigenvalue. This second highest eigenvalue has been connected with some internal structures of markets when dealing with assets from particular markets. For stock market indices, it has a different meaning that is peculiar to systems that operate at different times. Figures 16 to 21 show representations of the eigenvectors corresponding to the second highest eigenvalues for each of the intervals of time we are studying in this article. White rectangles correspond to positive values and dark rectangles correspond to negative values of the elements of the eigenvectors.

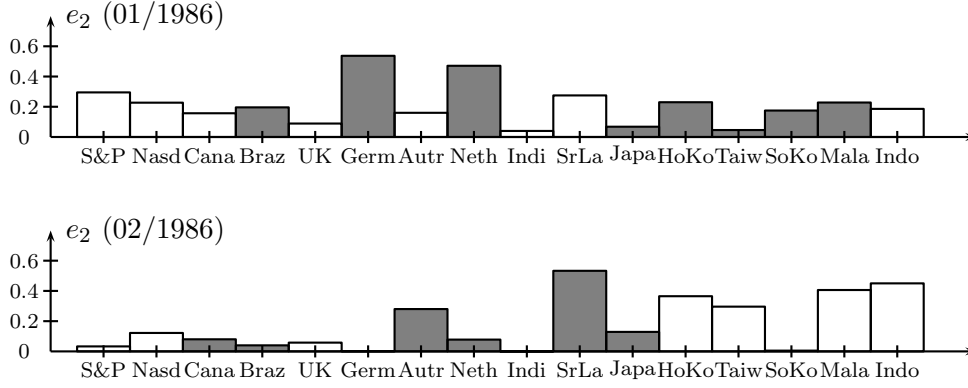


Fig. 16. Contributions of the stock market indices to eigenvector e_2 , corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the first and second semesters of 1986.

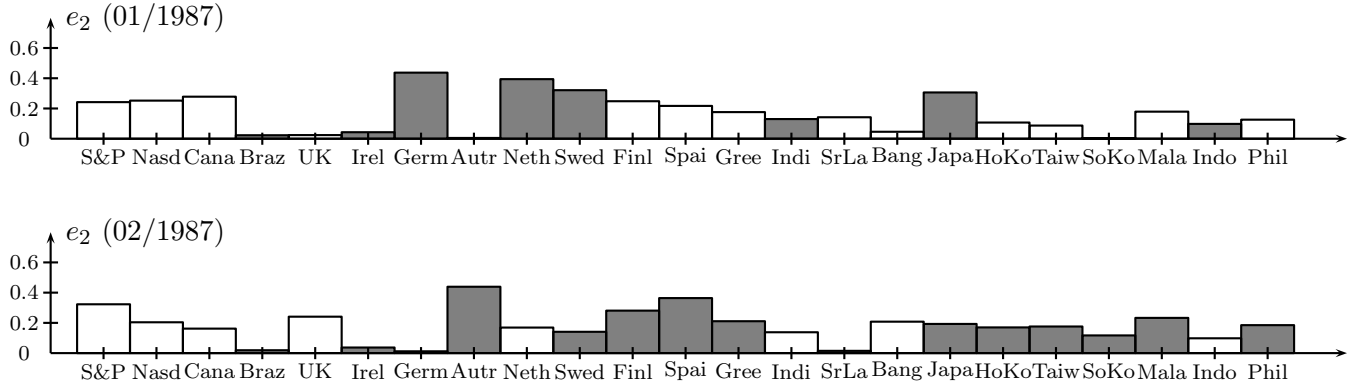


Fig. 17. Contributions of the stock market indices to eigenvector e_2 , corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the first and second semesters of 1987.

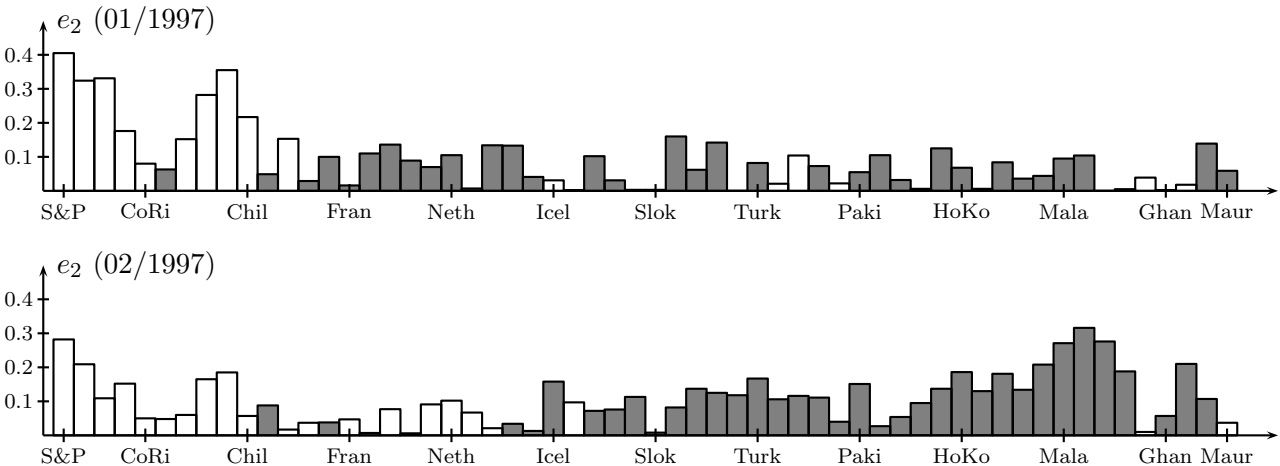


Fig. 18: contributions of the stock market indices to eigenvector e_2 , corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the first and second semesters of 1997. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, **CoRi**, Berm, Jama, Bra, Arg, **Chil**, Ven, Peru, UK, Irel, **Fran**, Germ, Swit, Austr, Belg, **Neth**, Swed, Denm, Finl, Norw, **Icel**, Spai, Port, Gree, CzRe, **Slok**, Hung, Pola, Esto, Russ, **Turk**, Isra, Leba, SaAr, Ohma, **Paki**, Indi, SrLa, Bang, Japa, **HoKo**, Chin, Taiw, SoKo, Thai, **Mala**, Indo, Phil, Aust, Moro, **Ghan**, Keny, SoAf, **Maur**.

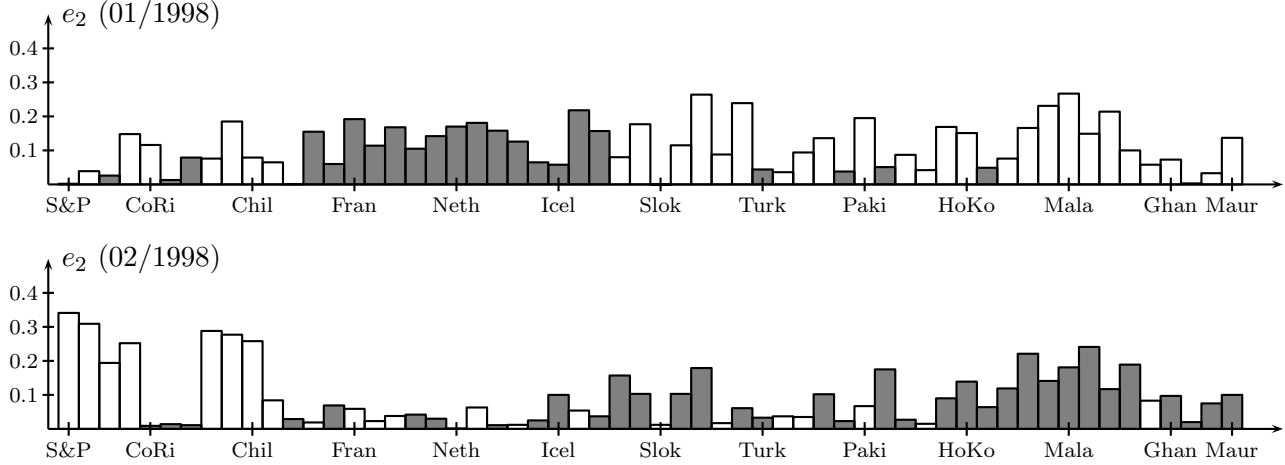


Fig. 19. Contributions of the stock market indices to eigenvector e_2 , corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the first and second semesters of 1997. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, **CoRi**, Berm, Jama, Bra, Arg, **Chil**, Ven, Peru, UK, Irel, **Fran**, Germ, Swit, Austr, Belg, **Neth**, Swed, Denm, Finl, Norw, **Icel**, Spai, Port, Gree, CzRe, **Slok**, Hung, Pola, Esto, Russ, **Turk**, Isra, Leba, SaAr, Ohma, **Paki**, Indi, SrLa, Bang, Japa, **HoKo**, Chin, Taiw, SoKo, Thai, **Mala**, Indo, Phil, Aust, Moro, **Ghan**, Keny, SoAf, **Maur**.

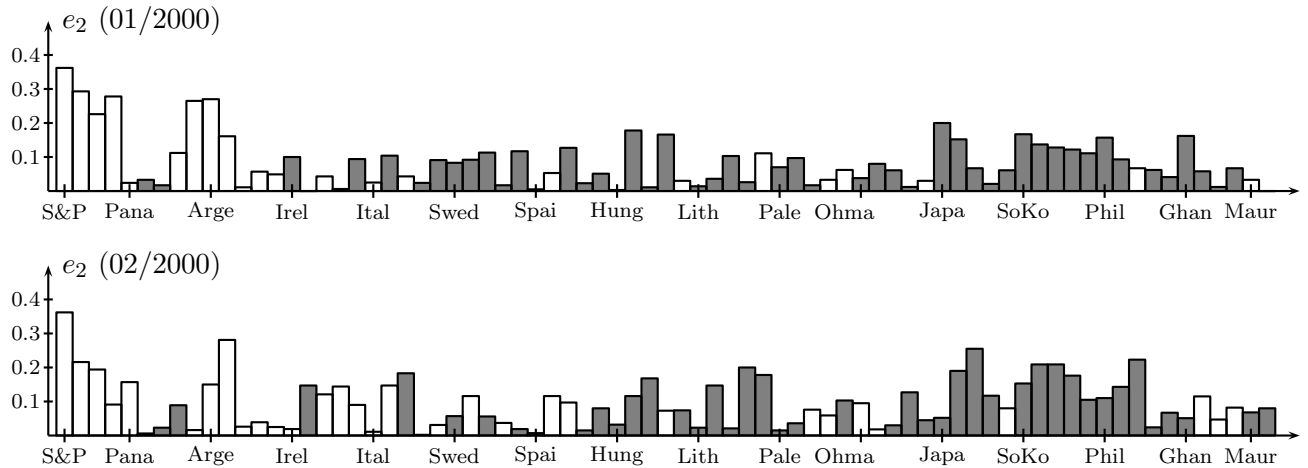


Fig. 20. Contributions of the stock market indices to eigenvector e_2 , corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the first and second semesters of 2000. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, **Pana**, CoRi, Berm, Jama, Braz, **Arge**, Chil, Vene, Peru, UK, **Irel**, Fran, Germ, Swit, Austr, **Ita**, Malt, Belg, Neth, Luxe, **Swed**, Denm, Finl, Norw, Icel, **Spai**, Port, Gree, CzRe, Slok, **Hung**, Pola, Roma, Esto, Latv, **Lith**, Ukra, Russ, Turk, Isra, **Pale**, Leba, Jord, SaAr, Qata, **Ohma**, Paki, Indi, SrLa, Bang, **Japa**, HoKo, Chin, Mong, Taiw, **SoKo**, Thai, Mala, Sing, Indo, **Phil**, Aust, Moro, Tuni, Egyp, **Ghan**, Nige, Keny, SoAf, **Maur**.

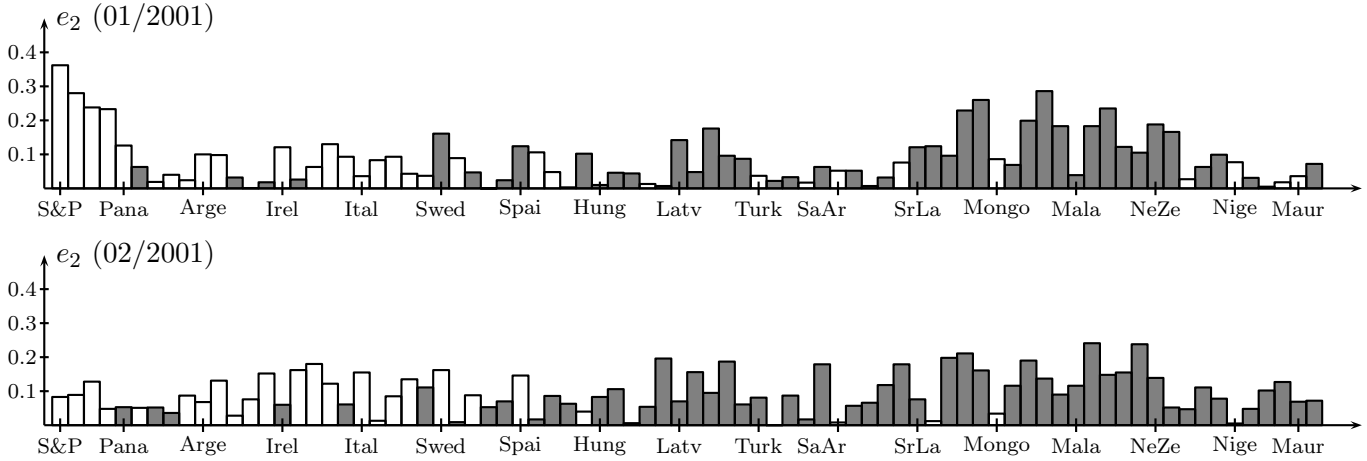


Fig. 21. Contributions of the stock market indices to eigenvector e_2 , corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the first and second semesters of 2001. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, **Pana**, CoRi, Berm, Jama, Braz, **Arge**, Chil, Vene, Peru, UK, **Irel**, Fran, Germ, Swit, Autr, **Ital**, Malt, Belg, Neth, Luxe, **Swed**, Denm, Finl, Norw, Icel, **Spai**, Port, Gree, CzRe, Slok, **Hung**, Pola, Roma, Bulg, Esto, **Latv**, Lith, Ukra, Russ, Kaza, **Turk**, Isra, Pale, Leba, Jord, **SaAr**, Qata, Ohma, Paki, Indi, **SrLa**, Bang, Japa, HoKo, Chin, **Mong**, Taiw, SoKo, Thai, Viet, **Mala**, Sing, Indo, Phil, Aust, **NeZe**, Moro, Tuni, Egypt, Ghan, **Nige**, Keny, Bots, SoAf, **Maur**.

What is not clear at the beginning, but becomes more evident in later times and with more data, is that there are two main blocks that move together as a second approximation to the market movement, and that those blocks are related with time zones, which reflect the operation hours of the stock exchanges. Usually, from Eastern Europe to Pacific Asia, indices belong to the second group, which appears with negative values in the eigenvectors. This is a characteristic peculiar to data related with international financial indices, and does not appear in ordinary correlations between assets in the same stock exchange. A certain contamination exists from other internal structures, like an European Union, quite clear in the data prior to the first semester of 1998, but further studies into the future (2007 to 2010) reveal that this separation in two blocks persists and even increases in time. A third highest eigenvalue sometimes also stand out of the noisy region, but contamination by random noise makes it very difficult to analyze any internal structure revealed by its corresponding eigenvector.

4 Conclusion

As we have seen in this article, correlation matrices between international stock market indices may be used in the construction of networks based on distance thresholds between the indices. By varying the thresholds, one may contemplate the cluster structure of those networks at different levels, revealing two strong and persistent clusters, one of American and the other of European indices, the formation of a Pacific Asian cluster in the 90's, and the slow integration of some of the indices. One could also see that those networks tend to shrink in size in times of crises, and, studying randomized data, that there are values of thresholds above which noise takes over. It was also shown that the eigenvectors of the correlation matrix hold some important information about the networks formed from it. In particular, the one corresponding to the second highest eigenvalue shows a structure based on the difference in operation hours between markets. Still, none of this study gives us information about causality between networks, and this shall indeed be a very interesting topic for future research.

Acknowledgements

The author acknowledges the support of this work by a grant from Insper, Instituto de Ensino e Pesquisa.

I am also grateful to Siew Ann Cheong, for useful discussions, and to the attenders and organizers of the Econophysics Colloquium 2010. This article was written using \LaTeX , all figures were made using PSTricks, and the calculations were made using Matlab and Excel. All data are freely available upon request on leonidassj@insper.edu.br.

References

- [1] R.N. Mantegna, *Hierarchical structure in financial markets*, The European Phys. J. B **11** (1999) 193.
- [2] G. Bonanno, F. Lillo, and R.N. Mantegna, *High-frequency cross-correlation in a set of stocks*, Quantitative Finance **1** (2001) 96-104.
- [3] S. Micchichè, G. Bonanno, F. Lillo, and R.N. Mantegna, *Degree stability of a minimum spanning tree of price return and volatility*, Physica A **324** (2003) 66-73.
- [4] G. Bonanno, G. Caldarelli, F. Lillo, S. Micchichè, N. Vandewalle, and R.N. Mantegna, *Networks of equities in financial markets*, The European Phys. J. B **38**, (2004) 363–371.
- [5] R. Coelho, C.G. Gilmore, B. Lucey, P. Richmond, and S. Hutzler, *The evolution of interdependence in world equity markets - evidence from minimum spanning trees*, Physica A **376** (2007) 455-466.
- [6] C. Borghesi, M. Marsili, and S. Micchichè, *Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode*, Phys. Rev. E **76** (2007), 026104.
- [7] J.G. Brida and W.A. Risso, *Multidimensional minimal spanning tree: the Dow Jones case*, Physica A **387** (2008) 5205-5210.
- [8] P. Siczka and J.A. Hołyst, *Correlations in commodity markets*, Physica A **388** (2009), 1621-1630.
- [9] J. Kwapien, S. Gworek, S. Drożdż, and A. Górski, *Analysis of a network structure of the foreign currency exchange market*, J. of Economic Interaction and Coordination **4** (2009), 55-72.
- [10] J. Kwapien, S. Gworek, and S. Drożdż, *Structure and evolution of the foreign exchange networks*, Acta Physica Polonica B **40** (2009), 175-194.
- [11] S. Drożdż, J. Kwapien, and J. Speth, *Coherent patterns in nuclei and in financial markets*, AIP Conf. Proc. **1261** (2010), 256-264.
- [12] D.J. Fenn, M.A. Porter, P.J. Mucha, M. McDonald, S. Williams, N.F. Johnson, and N.S. Jones, *Dynamical clustering of exchange rates*, preprint (2010) arXiv:0905.4912v2.
- [13] Y. Zhanga, G.H.T. Leea, J.C. Wonga, J.L. Kokb, M. Prustyb, and S.A. Cheong, *Will the US Economy Recover in 2010? A Minimal Spanning Tree Study*, (2010) Physica A **390** (2011) 2020-2050.
- [14] M. Keskin, B. Deviren, and Y. Kocakaplan, *Topology of the correlation networks among major currencies using hierarchical structure methods*, Physica A **390** (2011), 719-730.
- [15] L. Sandoval Jr., *Pruning a minimum spanning tree*, to appear in Physica A (2012).
- [16] C. Coronello, M. Tumminello, F. Lillo, S. Micchichè, and R.N. Mantegna, *Sector identification in a set of stock return time series traded at the London Stock Exchange*, Acta Phys. Pol. B **36** (2005) 2653-2679.
- [17] T. Aste and T. Di Matteo, *Correlation filtering in financial time series*, Noise and Fluctuations in Econophysics and Finance, Proceedings of the SPIE 5848 (2005) 100-109.
- [18] M. Tumminello, T. Di Matteo, T. Aste, and R.N. Mantegna, *Correlation based networks of equity returns sampled at different time horizons*, Eur. Phys. J. B **55** (2007) 209–217.

- [19] C. Coronello, M. Tumminello, F. Lillo, S. Micchichè, and R.N. Mantegna, *Economic sector identification in a set of stocks traded at the New York Stock Exchange: a comparative analysis*, Proceedings of the SPIE, vol. 6601, 66010T (2007).
- [20] M. Eryiğit and R. Eryiğit, *Network structure of cross-correlations among the world market indices*, Physica A **388** (2009), 3551-3562.
- [21] M. Tumminello, F. Lillo, and R.N. Mantegna, *Correlation, hierarchies, and networks in financial markets*, Journal of Economic Behavior & Organizations **75** (2010), 40-58.
- [22] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertész, *Dynamic asset trees and portfolio analysis*, The European Phys. J. B **30** (2002) 285-288.
- [23] J.-P. Onnela, A. Chakraborti, and K. Kaski, *Dynamics of market correlations: taxonomy and portfolio analysis*, Phys. Rev. E **68** (2003) 1-12.
- [24] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertész, *Dynamic asset trees and Black Monday*, Physica A **324** (2003) 247-252.
- [25] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Asset trees and asset graphs in financial markets*, Phys. Scripta T **106** (2003) 48-54.
- [26] J.-P. Onnela, K. Kaski, and J. Kertész, *Clustering and information in correlation based financial networks*, Eur. Phys. J. B **38** (2004) 353-362.
- [27] S. Sinha and R.K. Pan, *Uncovering the internal structure of the Indian financial market: cross-correlation behavior in the NSE*, in “Econophysics of markets and business networks”, Springer (2007) 215-226.
- [28] M. Ausloos and R. Lambiotte, *Clusters or networks of economies? A macroeconomy study through gross domestic product*, Physica A **382** (2007) 16-21.
- [29] L. Sandoval Jr., *A Map of the Brazilian Stock Market*, (2011) arXiv:1107.4146v1.
- [30] L. Sandoval Jr. and I. De P. Franca, *Correlation of financial markets in times of crisis*, Physica A **391**, (2012) 187-208.
- [31] D.Y. Kenett, Y. Shapira, and E. Ben-Jacob, *RMT assessments of the market latent information embedded in the stock’s raw, normalized, and partial correlations*, J. Prob. Stat. **2009**, (2009) 1-13.
- [32] D.Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R. Mantegna, and E. Ben-Jacob, *Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market*, Plos One **5**, (2010).
- [33] D.Y. Kenett, Y. Shapira, A. Madi, S. Bransburg-Zabary, G. Gur-Gershgoren, and E. Ben-Jacob, *Dynamics of stock market correlations*, AUCO Czech Economic Review **4**, (2010) 330-340.
- [34] Y. Shapira, D.Y. Kenett, and E. Ben-Jacob, *The index cohesive effect on stock market correlations*, Eur. Phys. J. B **72**, (2009) 657-669.
- [35] M. L. Mehta, *Random Matrices*, (2004) Academic Press.
- [36] V.A. Marčenko and L.A. Pastur, *USSR-Sb* **1** (1967) 457-483.